UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/696,714	10/29/2003	Simon S. Moy	NVDA/P000720	6459

26291 7590 11/17/2005

MOSER, PATTERSON & SHERIDAN L.L.P.
595 SHREWSBURY AVE, STE 100
FIRST FLOOR
SHREWSBURY, NJ 07702

EXAMINER	
FLOURNOY, HORACE L	

ART UNIT	PAPER NUMBER
2189	

DATE MAILED: 11/17/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/696,714	MOY ET AL.	
	Examiner	Art Unit	
	Horace L. Flournoy	2189	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *29 October 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-20* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-20* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *29 October 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

The instant application having Application No. **10/696,714** has a total of ***20*** claims pending in the application; there are ***3*** independent claims and ***17*** dependent claims, all of which are ready for examination by the examiner.

Information Concerning Oath/Declaration

Oath/Declaration

The applicant's oath/declaration has been reviewed by the examiner and is found to conform to the requirements prescribed in **37 C.F.R. 1.63**.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1-20 are rejected under 35 U.S.C. 102(b) as being anticipated by **Arimilli et al. (U.S. Patent No. 6,463,507 hereafter referred to as Arimilli).**

With respect to independent claim 1,

"A method for caching versions of data, comprising:

storing a first version of data in a first level 1 cache;" is disclosed in column 4, lines 34-35.

Arimilli discloses in **column 4, lines 34-35, "...the entire cache line of, say, 64 bytes must be loaded into the L1 cache..."**

"...storing a second version of data in a second level 1 cache; and..." is disclosed in columns 3, line 66- column 4, lines 1-12 and FIG. 4.

The examiner interprets this limitation as storing data in another L1 cache directory.

Arimilli teaches, for example, a set associative cache outlined in column 4, lines 1-12. A set associative cache can store a second version of data in a second level 1 cache or directory.

"...storing the first version of data in a level 2 cache." is disclosed in column 5, lines 28-33.

Arimilli discloses in **column 5, lines 28-33, "If inclusivity is maintained between the upper level and lower level caches, the determination that the copy of the target memory block is located in the upper level cache is**

accomplished by simply searching a directory of the lower level cache. The upper level cache is updated by the lower level cache."

With respect to claim 2,

"The method according to claim 1, further comprising invalidating the second level 1 cache when the second version of data is no longer being used by an execution unit." is disclosed in column 10, lines 53-59.

Arimilli discloses in **column 10, lines 53-59, "In the present invention, however, the invalidate request is guaranteed to occur, since L1 directory writes (invalidates) are controlled explicitly by the L2 cache, so no handshaking between the caches is necessary, and the L2 may immediately send a "clean" (or "null") response to the requesting device."**

Arimilli teaches invalidating (invalidates) multiple L1 caches (directory), therefore, Arimilli teaches invalidating a second level 1 cache when an execution unit is no longer using (requesting unit) a second version of data.

With respect to claim 3,

"The method according to claim 1, further comprising including a version tag with a read request to the level 2 cache." is disclosed in column 4, lines 44-53 and column 5, lines 15-19.

Arimilli discloses in **column 4, lines 44-53**, **"Write-back caches use state information bits to maintain consistency within the overall memory hierarchy (coherency), combined with the monitoring (snooping) of memory operations. One example of the state information is that supplied by the "MESI" cache coherency protocol, wherein a cache line can be in one of four coherency states: Modified, Exclusive, Shared or Invalid. Cache coherency protocols introduce further complexities and requirements into the interaction of the caches."** Arimilli further discloses in **column 5, lines 15-19**, **"...determining that a copy of the target memory block is located in an upper level cache associated with a second device of the computer system, and responding to the request using a lower level cache associated with the second device..."**

Arimilli teaches including a version tag (state information bit - MESI) and a read request (responding to the request using a lower level cache) to the level 2 cache (interaction of the caches).

With respect to claims 4, 12, and 13,

"The method according to claim 1, further comprising copying data from a location in the level 2 cache to a location in a level 2 cache backup." is disclosed in FIG. 4, elements 230, 232, and 240 and column 10, lines 17-30.

Arimilli discloses in **column 10, lines 17-26**, **"L2 cache 202 explicitly picks the victim and set for the reload, using an L1 least recently used (LRU) unit 228 and victim select logic 230 controlled by L2 controller 214. This approach has the added benefit of more easily maintaining inclusivity. It also moves the victimization process further away from the critical path. Moreover, since L2 controller 214 sees all of the L1 load addresses, it can maintain a hybrid L2 LRU 232."**

The examiner recognizes a level 2 cache backup as an L2 victim select (element 234), store cache (element 240), or L2 LRU (element 232), as taught by Arimilli.

With respect to claim 5,

"The method according to claim 4, further comprising associating a version tag with the data copied to the location in the level 2 cache backup." is disclosed as stated supra in claims 3 and 4 and in column 10, lines 61-67 – column 11, lines 1-2.

Arimilli discloses in **column 10, lines 61-67 – column 11, lines 1-2**, **"Still another advantage of the foregoing construction is that the L1 cache may be a store-through (write-through) cache, simplifying state information. Only one state bit is provided, a valid/invalid bit, unlike in traditional art such as the MESI protocol mentioned in the Background which uses two bits with a write-back cache. Using a store-through L1 cache, there are no**

cast outs or retry pushes at that level. All store operations issued by the core are forwarded to the L2 cache regardless of whether they hit or miss."

Arimilli teaches associating a version tag (MESI) with the data copied to the location in the level 2 cache backup. Note: Data copied to the L2 cache also has the capacity to be copied to the level 2 cache backup as outlined above in claim 4.

With respect to claim 6,

" The method according to claim 4, further comprising retiring the location in the level 2 cache backup when the data stored in the location is no longer being used by an execution unit." is disclosed in claim 2, in column 10, lines 15-60 and FIG. 4, element 232.

The examiner also notes that the L2 LRU (element 232) is a backup L2 cache that retires the data stored (using LRU protocol) in the location no longer being used by an execution unit.

With respect to claim 7,

"The method according to claim 1, further comprising copying at least a portion of data from the first level 1 cache to the second level 1 cache." is

disclosed in column 8, lines 35-37, column 13, lines 40-47 and FIG. 4 elements 228 and 230.

Arimilli discloses in **column 8, lines 35-37**, **"Any request for a load operation is sent along request bus 212 to L1 data directory 208 and L1 data entry array 210." Arimilli further discloses in column 13, lines 40-47, "For example, while the illustrative embodiment provides only vertical L1 and L2 cache levels for a single associated processor, the invention can be extended to additional cache levels as well, or to multi-processor systems, or to cache hierarchies having vertical caches that support a processor core cluster."**

Arimilli teaches that FIG. 4 can be replicated to form more L1 and L2 caches.

With respect to claim 8,

"The method according to claim 1, further comprising copying at least a portion of data from the first level 1 cache to the level 2 cache." is disclosed in column 5, lines 28-33.

Arimilli discloses in **column 5, lines 28-33**, **"If inclusivity is maintained between the upper level and lower level caches, the determination that the copy of the target memory block is located in the upper level cache is accomplished by simply searching a directory of the lower level cache. The upper level cache is updated by the lower level cache."**

With respect to claim 9,

"The method according to claim 1, further comprising updating a version tag in the first level 1 cache when data is stored in the second level 1 cache and not stored in the first level 1 cache." is disclosed in column 4, lines 44-53 and column 5, lines 15-19 and in column 10, lines 61-67 – column 11, lines 1-2.

The examiner interprets this claim as analogous to the Modified, Exclusive, Shared or Invalid cache coherency protocol.

Arimilli discloses in **column 4, lines 44-53, "Write-back caches use state information bits to maintain consistency within the overall memory hierarchy (coherency), combined with the monitoring (snooping) of memory operations. One example of the state information is that supplied by the "MESI" cache coherency protocol, wherein a cache line can be in one of four coherency states: Modified, Exclusive, Shared or Invalid. Cache coherency protocols introduce further complexities and requirements into the interaction of the caches."** Arimilli further discloses in **column 5, lines 15-19, "...determining that a copy of the target memory block is located in an upper level cache associated with a second device of the computer system, and responding to the request using a lower level cache associated with the second device..."** Arimilli also, discloses in **column 10, lines 61-67 – column 11, lines 1-2, "Still another advantage of the**

foregoing construction is that the L1 cache may be a store-through (write-through) cache, simplifying state information. Only one state bit is provided, a valid/invalid bit, unlike in traditional art such as the MESI protocol mentioned in the Background which uses two bits with a write-back cache. Using a store-through L1 cache, there are no cast outs or retry pushes at that level. All store operations issued by the core are forwarded to the L2 cache regardless of whether they hit or miss."

With respect to claim 10,

"The method according to claim 1, further comprising: storing the second version of data in the level 2 cache;" is disclosed in column 8, lines 43-57 and FIG. 4.

Arimilli discloses in **column 8, lines 53-57**, **"Instead of load queues, the present invention handles L1 misses by passing down other information to L2 cache 200, which allows for the later placement of the requested data in the appropriate register rename 206."**

"...not storing the second version of data in the first level 1 cache; and..." is disclosed in column 9, lines 52-57.

Arimilli discloses in **column 9, lines 52-57**, **"First, L2 cache 202 may choose to not reload L1 cache 200. Second, L1 cache 200 may refuse to accept a reload that has been proffered from L2 cache 202. Generally, the**

reload will occur whenever a determination is made that it would be efficient to currently load the cache line into the upper level."

"...marking the first level 1 cache as invalid." is disclosed as stated supra in claim 2 and 9.

With respect to independent claim 11,

"A streaming processing array, comprising: a first execution unit configured to process data and including a first level 1 cache;" is disclosed in column 9, lines 19-24, lines 30-35, FIG. 1, and FIG. 4, elements 206, 210 and 218.

Arimilli discloses in **column 9, lines 19-24, "Thus, upon an L1 cache miss, if the requested data is present in L2 cache 202, it can be directly ported to register rename 206 via register bus 224 without first having to wait for the entire cache line to be received by the L1 entry array."** Arimilli further discloses in **column 9, lines 30-35, "In this manner, even if the core is executing an excessive number of sequential or nearly sequential load operations, there is no stalling as the load request is simply piped downstream, and then the data is piped back up to the register renames."**

Arimilli teaches a streaming processing array as shown in the references above (see also FIG. 4 elements 206, 210, and 218). Arimilli also teaches a first execution unit (FIG. 1) that can process data and includes a first level 1 cache (as shown in FIG. 4).

[Note: the limitation "streaming processing array" is henceforth interpreted as stated supra in the following claims 12-20].

"...a second execution unit configured to process data and including a second a level 1 cache; and..." is disclosed in column 13, lines 40-47.

Arimilli further discloses in **column 13, lines 40-47, "For example, while the illustrative embodiment provides only vertical L1 and L2 cache levels for a single associated processor, the invention can be extended to additional cache levels as well, or to multi-processor systems, or to cache hierarchies having vertical caches that support a processor core cluster."**

Arimilli teaches that FIGs. 1 and 4 can be replicated to form more execution units and hence more L1 and L2 caches.

"...a level 2 cache coupled to both the first execution unit and the second execution unit." is disclosed as stated supra and in FIG. 1, elements 18 and 20 and FIG. 4 elements 105.

With respect to claim 12,

"The streaming processing array of claim 11, further comprising a level 2 cache backup coupled to the level 2 cache." is disclosed as stated supra in claim 4.

With respect to claim 13,

"The streaming processing array of claim 12, wherein the level 2 cache is configured to output data to the level 2 cache backup." is disclosed as stated supra in claim 4.

With respect to claim 14,

"The streaming processing array of claim 12, further comprising a controller configured to associate a version tag with data stored in each location in the level 2 cache backup." is disclosed in column 9, lines 27-30, column 10, lines 4-6 and FIG. 4, element 214.

Arimilli discloses in **column 9, lines 27-30, "In order to allow the L2 controller to determine whether it needs to honor the request, a flag is provided to L2 controller 214 from L1 data directory 208 to indicate the hit/miss status."**

Arimilli teaches the streaming processing array of claim 12 (stated supra in claim 13), further comprising a controller (L2 controller 214) configured to associate a version tag (hit/miss status) with data stored in each location (data directory 208) in the level 2 cache backup (also see FIG. 4 element 216).

With respect to claim 15,

"The streaming processing array of claim 14, wherein the controller is configured to retire a location in the level 2 cache backup when a version

tag associated with the data stored in the location is no longer being used by an execution unit." is disclosed in column 10, lines 17-26.

Arimilli discloses in **column 10, lines 17-26, "L2 cache 202 explicitly picks the victim and set for the reload, using an L1 least recently used (LRU) unit 228 and victim select logic 230 controlled by L2 controller 214. This approach has the added benefit of more easily maintaining inclusivity. It also moves the victimization process further away from the critical path. Moreover, since L2 controller 214 sees all of the L1 load addresses, it can maintain a hybrid L2 LRU 232."**

Arimilli teaches the L2 cache controller (FIG. 4, element 214), which controls the L2 LRU and victim select logic, and therefore can retire a location in the level 2 cache backup (L2 LRU/ VICTIM SELECT) when a version tag associated with the data stored in the location is no longer being used by an execution unit (Invalid – MESI).

With respect to claim 16,

"The streaming processing array of claim 11, wherein the first level 1 cache is configured to output data to and to receive data from the second level 1 cache." is disclosed as stated supra in claim 10 and in column 13, lines 40-47.

Arimilli further discloses in **column 13, lines 40-47, "For example, while the illustrative embodiment provides only vertical L1 and L2 cache levels**

for a single associated processor, the invention can be extended to additional cache levels as well, or to multi-processor systems, or to cache hierarchies having vertical caches that support a processor core cluster."

Arimilli teaches a multi-processor system (separate L1 caches) which can output data and receive data from each other (stated supra).x

With respect to claim 17,

"The streaming processing array of claim 11, wherein the first level 1 cache is configured to output data to the level 2 cache." is disclosed as stated supra in claim 8.

With respect to claim 18,

"The streaming processing array of claim 11, wherein the streaming processing array resides within a programmable graphics processor." is disclosed in column 7, lines 51-59, and FIGs. 1-3.

Arimilli discloses in **column 7, lines 51-59**, **"Display 196, which is controlled by display controller 198, is used to display visual output generated by data processing system 120. Such visual output may include text, graphics, animated graphics, and video. Display 196 may be implemented with CRT-based video display, an LCD-based flat panel display, or a gas plasma-based flat-panel display. Display controller 198**

includes electronic components required to generate a video signal that is sent to display 196."

Arimilli discloses a programmable graphics processor (data processing system 120). Furthermore, the graphics processing system can be programmed through the use of code via the CD-ROM device (element 178). Also, FIG. 3 teaches a CPU (element 150) which is a graphics processor.

With respect to claim 19,

"The streaming processing array of claim 18, wherein the programmable graphics processor is coupled to a host computer." is disclosed in FIG. 3.

Note: see elements 150 and 105; and also FIG. 2, element 105.

With respect to claim 20,

"A system for processing data, the system comprising:

means for storing a first version of data in a first level 1 cache;

means for storing a second version of data in a second level 1 cache; and

means for storing the first version or the second version of data in a level 2 cache."

Claim 20 is interpreted under 35 U.S.C. 112, 6th paragraph.

The Court of Appeals for the Federal Circuit, in its en banc decision In re Donaldson Co., 16 F.3d 1189, 29 USPQ2d 1845 (Fed. Cir. 1994), decided that a "means-or-step-plus-function"

limitation should be interpreted in a manner different than patent examining practice had previously dictated. The Donaldson decision affects only the manner in which the scope of a "means or step plus function" limitation in accordance with 35 U.S.C. 112, sixth paragraph, is interpreted during examination. Donaldson does not directly affect the manner in which any other section of the patent statutes is interpreted or applied.

When making a determination of patentability under 35 U.S.C. 102 or 103, past practice was to interpret a "means or step plus function" limitation by giving it the "broadest reasonable interpretation." Under the PTO's long-standing practice this meant interpreting such a limitation as reading on any prior art means or step which performed the function specified in the claim without regard for whether the prior art means or step was equivalent to the corresponding structure, material or acts described in the specification. However, in Donaldson, the Federal Circuit stated:

> Per our holding, the "broadest reasonable interpretation" that an examiner may give means-plus-function language is that statutorily mandated in paragraph six. Accordingly, the PTO may not disregard the structure disclosed in the specification corresponding to such language when rendering a patentability determination. (MPEP 2181)

Accordingly, the Examiner notes that the means or system/structure for practice of the invention disclosed in Claim 20, of applicant's specification **(as evaluated in paragraphs [0027]-[0029])** is taught in the previous rejection of claim 1, as stated in the present action.

CONCLUSION

Status of Claims in the Application

The following is a summary of the treatment and status of all claims in the application as recommended by **M.P.E.P. 707.07(i):**

Claims rejected in the Application

Per the instant office action, claims **1-20** have received a first action on the merits and are subject of a first action non-final.

Direction of Future Correspondences

Any inquiry concerning this communication or earlier communication from the examiner should be directed to Horace L. Flournoy whose telephone number is (571) 272-2705. The examiner can normally be reached on Monday through Friday 8:00 AM to 5:30 PM (ET).

Important Note

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Don Sparks can be reached on (571) 272-4201. The fax phone numbers for the organization where this application or proceeding is assigned is (703) 746-7239.

Information regarding the status of an Application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or PUBLIC PAIR. Status information for unpublished applications is available through Private Pair only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (571) 272-2100.

DONALD SPARKS
SUPERVISORY PATENT EXAMINER

Horace L. Flournoy

Patent Examiner

Art unit: 2189

Primary Patent Examiner

Technology Center 2100